Accuracy of automated liver contouring, fat fraction, and R2* measurement on gradient multiecho magnetic resonance images

Stocker, Daniel ; Bashir, Mustafa R ; Kannengiesser, Stephan A R ; Reiner, Cäcilia S

Abstract: OBJECTIVE: This study aimed to evaluate the performance of an automated workflow of volumetric liver proton density fat fraction (PDFFvol) and R2* quantification with automated inline liver volume (LV) segmentation. METHODS: Dual-echo and multiecho Dixon magnetic resonance images were evaluated in 74 consecutive patients (group A, PDFF < 10%; B, PDFF 10%; C, R2* 100 s; D, post-hemihepatectomy). The values of PDFFvol and R2*vol measurements across the LV were generated on multiecho images in an automated fashion based on inline liver segmentation on dual-echo images. Similar measurements were performed manually. RESULTS: Using the inline algorithm, the mis-segmented LV was highest in group D (80%). There were no significant differences between automated and manual measurements of PDFFvol. Automated R2*vol was significantly lower than manual R2*vol in group A (P = 0.004). CONCLUSIONS: Inline LV segmentation performed well in patients without and with hepatic steatosis. In cases with iron overload and post-hemihepatectomy, extrahepatic areas were erroneously included to a greater extent, with a tendency toward overestimation of PDFFvol.

DOI: https://doi.org/10.1097/RCT.0000000000000759

Accuracy of Automated Liver Contouring, Fat Fraction, and R2* Measurement on Gradient Multiecho Magnetic Resonance Images

Daniel Stocker, MD, Mustafa R. Bashir, MD,†‡*
*Stephan A.R. Kannengiesser, PhD,§ and Cäcilia S. Reiner, MD**

Objective: This study aimed to evaluate the performance of an automated workflow of volumetric liver proton density fat fraction ($PDFF_{vol}$) and R2* quantification with automated inline liver volume (LV) segmentation.

Methods: Dual-echo and multiecho Dixon magnetic resonance images were evaluated in 74 consecutive patients (group A, PDFF < 10%; B, PDFF ≥ 10%; C, R2* ≥ 100 s^{-1}; D, post-hemihepatectomy). The values of $PDFF_{vol}$ and $R2^*_{vol}$ measurements across the LV were generated on multiecho images in an automated fashion based on inline liver segmentation on dual-echo images. Similar measurements were performed manually.

Results: Using the inline algorithm, the mis-segmented LV was highest in group D (80%). There were no significant differences between automated and manual measurements of $PDFF_{vol}$. Automated $R2^*_{vol}$ was significantly lower than manual $R2^*_{vol}$ in group A ($P = 0.004$).

Conclusions: Inline LV segmentation performed well in patients without and with hepatic steatosis. In cases with iron overload and post-hemihepatectomy, extrahepatic areas were erroneously included to a greater extent, with a tendency toward overestimation of $PDFF_{vol}$.

Key Words: liver fat, liver iron, dixon technique

(*J Comput Assist Tomogr* 2018;00: 00–00)

Magnetic resonance imaging (MRI) is a noninvasive tool for assessment of diffuse liver deposition disease, and it accurately assesses hepatic fat and iron deposition.[1–5] Nowadays, an efficient way for simultaneous fat and iron quantification is based on 3-dimensional (3D) multigradient-echo imaging.[6] To allow for simultaneous quantification of hepatic fat and iron deposition and correction of other confounding factors, image reconstructions such as the multistep adaptive fitting algorithm have been developed, which incorporate T2* and other corrections.[7,8]

The MRI technique for quantification of liver fat and iron deposition has clear advantages over the current standard method of liver parenchyma assessment, namely, percutaneous liver biopsy, which carries the potential of complications such as infection and bleeding.[9,10] Furthermore, with a liver biopsy, only a small part of the liver is evaluated, which may be a source of sampling error, especially in liver steatosis known to be inhomogeneous throughout the liver.[10–12] Techniques based on MRI, on the other hand, are noninvasive and provide information on the whole liver. However, the quantification of fat and iron deposition in the whole liver volume (LV) is time consuming because, first, an additional magnetic resonance (MR) sequence needs to be scanned and, second, multiple regions of interest (ROIs) or volumes of interest (VOIs) must be drawn manually to ensure analysis of a representative portion of the liver. To facilitate the MRI-based liver fat and iron quantification, a dedicated workflow has been proposed.[13] In this workflow, the liver is screened for fat and iron deposition with a T1-weighted dual-gradient-echo MR sequence with 2-point Dixon technique, which is part of routine liver MRI protocols.[14] This technique includes an algorithm for inline segmentation of the liver in water-only 3D image volumes from the T1-weighted dual-gradient-echo MR sequence. If the screening result is positive for fat and/or iron deposition, the additional multigradient-echo sequence for quantification is acquired and multiple ROIs need to be placed manually in the liver. To avoid time-consuming manual placement of multiple ROIs, measurements using automated sampling of the liver parenchyma would be desirable.

In the first commercially available version of a liver fat and iron quantification workflow, an automated propagation of the morphology-based inline liver segmentation result to the results from the 3D multigradient-echo MRI sequence with multistep adaptive fitting has been integrate for automated measurements of proton density fat fraction (PDFF) and transverse relaxation rate (R2*).

The purpose of this study was to evaluate how this workflow of liver fat and iron quantification with automated inline LV segmentation performs in patients with different degrees of liver fat and/or iron deposition disease and how the inline LV segmentation influences automated volumetric measurements of liver PDFF and R2* from the separate multiecho sequence compared with manual measurements.

MATERIALS AND METHODS

Study Population

Institutional review board approval was obtained for this retrospective analysis. Informed consent was waived for patients before 2016. From 2016 onward, patients had to give written informed consent for inclusion of their imaging data in future retrospective research.

The institutional Radiology Information System was searched for gadolinium ethoxybenzyl-diethylenetriaminepentaacetic acid–enhanced liver MRI examinations performed between December 2014 and September 2016 at our institution. A total of 598 examinations from unique patients (39.8% female [238/598]; mean age, 56 years [range, 16–89 years]) were identified, and their first available scans were reviewed by one radiologist (D.S.; 3 years of experience

From the *Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Zurich, Switzerland; †Department of Radiology and ‡Center for Advanced Magnetic Resonance Development, Duke University Medical Center, Durham, NC; and §MR Application Predevelopment, Siemens Healthcare, Erlangen, Germany.

Received for publication December 2, 2017; accepted April 2, 2018.

Correspondence to: Cäcilia S. Reiner, MD, Institute of Interventional and Diagnostic Radiology, University Hospital Zurich and University of Zurich, Raemistrasse 100, 8091 Zurich, Switzerland (e-mail: caecilia.reiner@gmx.at).

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Stephan A.R. Kannengiesser is an employee of Siemens Healthcare, Erlangen, Germany.

The authors declare no conflict of interest.

DOI: 10.1097/RCT.0000000000000759

in abdominal cross-section imaging). Patients were excluded for the following reasons: did not provide written informed consent (49.8%; 298/598), swap artifacts on 3D gradient dual-echo or multiecho Dixon sequences (4.7%; 28/598), age less than 18 years (0.7%; 4/598), and incomplete MRI protocol (2.9%; 19/598). Of the remaining 249 patients, consecutive patients without prior major liver resection were added into the following 3 subgroups, with a maximum group size of 30 patients: group A, PDFF < 10%; group B, PDFF ≥ 10%; and group C, R2* ≥ 100 s^{-1}. The PDFF and R2* results of manually drawn single ROIs from clinical reports were used to group patients. We included all patients who qualified for the specific subgroup into group A, B, or C to a maximum group size of 30 patients To evaluate the influence of liver shape on the inline segmentation algorithm, 10 consecutive patients with prior hemihepatectomy were included in group D that had not been included in group A, B, or C.

As a result, 74 consecutive patients (31.1% female [23/74]; mean age, 58 years [range, 22–82 years]) were identified: with 30 patients in group A (33.3% female [10/30]; mean age, 59 years [range, 22–82 years]), 14 patients in group B (28.6% female [4/14]; mean age, 58 years [range, 22–80 years]), 20 patients in group C (25% female [5/20]; mean age, 54 years [range, 31–76 years]), and 10 patients in group D (40% female [4/10]; mean age, 65 years [range, 51–73 years]). If the patients had a R2* ≥ 100 s^{-1}, they were always included into group C, whether the PDFF was less than 10% (in 16/20 patients) or at least 10% (in 4/20 patients). No patient in group A or B had R2* measurements greater than or equal to 100 s^{-1}. Patient demographics are summarized in Table 1, and a flowchart depicting the selection of the final study cohort can be found in Figure 1.

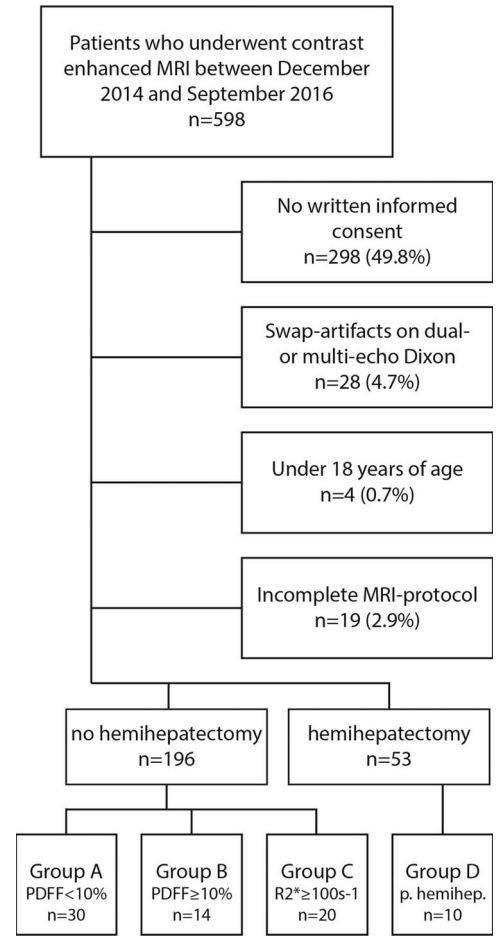

FIGURE 1. Flowchart of the study population. Note: The maximum group size was set at 30 for groups A to C and at 10 for group D.

Magnetic Resonance Imaging

Image acquisition was performed on 1 of 2 identical 3-T MRI scanners (MAGNETOM Skyra; Siemens Healthcare, Erlangen, Germany) using one flexible 18-channel body matrix coil combined with the integrated 32-channel spine matrix coil. All examinations included the following: (1) axial 3D T1-weighted dual-gradient-echo volume-interpolated breath-hold examination (VIBE) sequence with 2-point Dixon reconstruction (repetition time, 4.11 milliseconds; echo time, 1.31, 2.54 milliseconds; flip angle, 9 degrees; field of view, 380 mm; matrix, 320 × 217; slice thickness, 3 mm; voxel size, 1.2 × 1.2 × 3.0 mm^3) and (2) axial 3D multigradient-echo VIBE sequence with multistep adaptive fitting algorithm reconstruction to derive PDFF and R_2* maps (repetition time, 9.0 milliseconds; echo

TABLE 1. Patient Demographics

	Total (n = 74)	Group A (n = 30)	Group B (n = 14)	Group C (n = 20)	Group D (n = 10)
General features					
Male/female	51/23 (68.9%/31.1%)	20/10 (66.7%/33.3%)	10/4 (71.4%/28.6%)	15/5 (75%/25%)	6/4 (60%/40%)
Age, y	58 (22–82)	59 (22–82)	58 (22–80)	54 (31–76)	65 (51–73)
Height, cm	172.2 (155–189)	170.3 (156–182)	174.0 (155–187)	175.5 (160–189)	168.7 (155–187)
Weight, kg	79.9 (45–133)	72.9 (45–133)	85.6 (51–120)	89.7 (62–128)	76.1 (50–109)
BMI, kg/m^2	26.8 (18–41)	24.9 (18–41)	28.2 (20–40)	29.1 (23–40)	26.5 (20.3-31.2)
Mean PDFF, %	7 (0–28.6)	3.1 (0.3–9.8)	15.7 (10.5–28.6)	6.1 (0–24)	2.9 (0.6–9.4)
Mean R2*, s^{-1}	96.6 (1.4–427)	56 (26.6–98.5)	79.9 (57–99.1)	194.5 (101–427)	55.8 (1.4–104.8)
Indication for MRI					
Focal liver lesion	45 (60.8%)	23 (76.7%)	8 (57.1%)	4 (20%)	10 (100%)
Hemochromatosis	23 (31.1%)	1 (3.3%)	6 (42.9%)	16 (80%)	0
Cirrhosis	4 (5.4%)	4 (13.3%)	0	0	0
Others	2 (2.7%)	2 (6.7%)	0	0	0

Range is shown in parenthesis. Mean PDFF and R2* measurements of manually drawn single ROIs from clinical reports are presented.
Group A, PDFF < 10%; group B, PDFF ≥ 10%; group C, R2* ≥ 100 s^{-1}; group D, status post-hemihepatectomy.

time, 1.05, 2.46, 3.69, 4.92, 6.15, 7.38 milliseconds; flip angle, 4 degrees; field of view, 450 mm; matrix, 160 × 111; slice thickness, 3.5 mm; voxel size, 1.4 × 1.4 × 3.5 mm^3).

Automated volumetric liver PDFF and $R2^*_{vol}$ Measurements

Liver volume segmentation had been performed inline during the MRI scan on the water-only images of the dual-echo sequence using a learning-based liver segmentation algorithm,[15–17] and maps of the inline-segmented volume were available as a series of DICOM images. Intrahepatic vessels and hilar structures were not considered. Subsequently, this 3-dimensional liver VOI had been propagated inline to the results of the multiecho sequence by the scanner and was also available as a series of DICOM images. The values of volumetric liver PDFF ($PDFF_{vol}$) and of R2* ($R2^*_{vol}$) had been measured inline on the MRI system using this propagated VOI, and those results were available as tabulated results in a DICOM image.

For evaluation of the quality of inline LV segmentation, it was noted whether the selected volume was in part located outside the liver on the multiecho results. The volume of erroneously selected parts outside the liver was then manually measured using volumetry software (OsiriX v.4.0 32-bit; Pixmeo Sarl, Geneva, Switzerland).

Manual $PDFF_{vol}$ and $R2^*_{vol}$ Measurements

One radiology resident with 3 years of experience in abdominal cross-sectional imaging, blinded to the results of inline liver segmentation, manually segmented the whole liver parenchyma on the water image of the dual-echo sequence using volumetry software (Myrian version 1.21.1; Intrasense, Paris, France). The contour was delineated every few slices, depending on the contour changes compared with the previous drawn contour. The contour on the remaining slices was interpolated automatically; major fissures, the liver hilum, and the gallbladder fossa were excluded. The manually segmented VOI was propagated to the PDFF and R2* maps from the multiecho sequence. In case of erroneous VOI propagation (ie, due to different breath-hold), the VOI was manually adjusted on multiecho images. Mean values of $PDFF_{vol}$ and $R2^*_{vol}$ within the VOI were recorded.

Statistical Analysis

Related-samples Wilcoxon signed-rank tests were used to assess for significant differences between automated and manual values for LV, $PDFF_{vol}$, and $R2^*_{vol}$ for all groups. Bland-Altman plots were generated to assess differences between automated and manual measurements. Linear regression analysis was performed to evaluated correlation between automated and manual values for $PDFF_{vol,}$ and $R2^*_{vol}$ for all groups. Scatter plots were generated to visualize correlation between automated and manual values. Furthermore, absolute and relative differences for LV, $PDFF_{vol}$, and $R2^*_{vol}$ were calculated. Quantitative results are reported as mean ± SD and ranges. All statistical analyses were performed using SPSS (IBM SPSS Statistics 22; SPSS Inc, Chicago, IL) and Excel software (Microsoft Corporation, Redmond, WA). $P < 0.05$ was considered statistically significant.

RESULTS

A summary of the automated and manual measurement results for, $PDFF_{vol}$, $R2^*_{vol}$, and LV in groups A to D is provided in Table 2, and that for patients where the sample volume was located partly outside the liver is shown in Table 3. Case-by-case differences between automated and manual $PDFF_{vol}$ and $R2^*_{vol}$ measurements are shown in Figure 2. Results from linear regression analysis for automated versus manual $PDFF_{vol}$ and $R2^*_{vol}$ in groups A to D and for patients where the sample volume was located partly outside the liver are provided in Table 4. Scatter plots for $PDFF_{vol}$ and $R2^*_{vol}$ in groups A to D are shown in Figure 3 and Figure 4.

Group A

With inline liver segmentation, in 11 (36.7%) of 30 cases, the sampled volume was located partly outside the liver on multiecho results. The mean volume located outside the liver was

TABLE 2. Results of Automated and Manual Measurements of LV, $PDFF_{vol}$, and $R2^*_{vol}$ of Each Group

	Automated Measurements	Manual Measurements	Absolute Difference	*P*
Group A (n = 30)				
$PDFF_{vol}$, %	5.8 ± 4.3	6.2 ± 3.8	−0.3 ± 2.1 (−5.0 to 5.2)	0.154
$R2^*_{vol}$, s^{-1}	77.3 ± 64.8	82.4 ± 84.3	−5.1 ± 22.7 (−123.4 to 13.8)	0.004
LV, mL	1550.5 ± 416.7	1788.7 ± 754.0	−238.2 ± 623.1 (−3458 to 176)	0.0001
Group B (n = 14)				
$PDFF_{vol}$, %	16.3 ± 5.2	15.6 ± 5.1	0.6 ± 2.5 (−1.9 to 7.3)	0.843
$R2^*_{vol}$, s^{-1}	79.7 ± 20.3	79.2 ± 20.6	0.5 ± 3.8 (−2.9 to 13.1)	0.358
LV, mL	2143.6 ± 654.8	2344.3 ± 978.7	−200.7 ± 527.5 (−2014 to 34)	0.004
Group C (n = 20)				
$PDFF_{vo}$l, %	10.6 ± 11.1	8.7 ± 5.5	1.8 ± 8.1 (−2.9 to 27.0)	0.330
$R2^*_{vol}$, s^{-1}	170.7 ± 81.2	195.3 ± 115.1	−24.6 ± 78.8 (−313.0 to 19.3)	0.123
LV, mL	1689.1 ± 348.2	1817.9 ± 394.3	−128.8 ± 139.9 (−451 to 152)	0.0005
Group D (n = 10)				
$PDFF_{vol}$, %	7.5 ± 6.3	4.4 ± 2.0	3.1 ± 5.4 (−1.1 to 13.9)	0.322
$R2^*_{vol}$, s^{-1}	64.4 ± 16.6	61.8 ± 11.5	2.6 ± 8.3 (−8.1 to 20.2)	0.375
LV, mL	1582.8 ± 318.4	1470.1 ± 284.3	112.7 ± 262.8 (−192 to 654)	0.432

Numbers are given as mean ± SD; numbers in parenthesis are ranges.

Group A, PDFF < 10%; group B, PDFF ≥ 10%; group C, R2* ≥ 100 s^{-1}; group D, status post-hemihepatectomy.

TABLE 3. Results of Automated and Manual Measurements of LV, $PDFF_{vol}$, and $R2^*_{vol}$ of Each Group in Patients Where the Sample Volume Was Located Partly Outside the Liver

	Volume Located Outside the Liver (Absolute)	Volume Located Outside the Liver (Relative)	Automated Measurements	Manual Measurements	Absolute Difference	*P*
Group A (n = 11/30; 37%)						
$PDFF_{vol}$, %	—	—	8.2 ± 5.4	7.5 ± 4.2	−0.7 ± 2.7 (−5.2 to 5)	0.413
$R2^*_{vol}$, s^{-1}	—	—	93.3 ± 89.7	104.8 ± 123.9	11.5 ± 35.7 (−13.8 to 123.4)	0.175
LV, mL	175.8 ± 202.6	11.5% ± 11.4%	1503.3 ± 386.7	1524.4 ± 377.8	21.1 ± 122.8 (−176 to 198)	0.638
Group B (n = 5/14; 35%)						
$PDFF_{vol}$, %	—	—	17.7 ± 5.1	15.3 ± 5.5	−2.4 ± 3.3 (−7.3 to 1)	0.438
$R2^*_{vol}$, s^{-1}	—	—	90 ± 23.9	87.8 ± 25.9	−2.3 ± 5.4 (−13.1 to 1.6)	1
LV, mL	232.9 ± 152.9	11.9% ± 7.5%	1903.6 ± 274.4	1969.8 ± 336.2	66.2 ± 107 (−34 to 270)	0.313
Group C (n = 7/20; 35%)						
$PDFF_{vol}$, %	—	—	19 ± 14.2	11.7 ± 6.3	−7.2 ± 11.4 (−27 to 1.5)	0.398
$R2^*_{vol}$, s^{-1}	—	—	173.6 ± 49.9	244.3 ± 131.8	70.7 ± 116.2 (−1.5 to 313)	0.091
LV, mL	436 ± 555.3	27.7% ± 36.2%	1664.9 ± 340.6	1820.3 ± 372	155.4 ± 195.2 (−152 to 451)	0.109
Group D (n = 8/10; 80%)						
$PDFF_{vol}$, %	—	—	8.3 ± 6.3	4.4 ± 1.5	−3.9 ± 5.5 (−13.9 to 1.1)	0.313
$R2^*_{vol}$, s^{-1}	—	—	66.7 ± 16.2	62.7 ± 11.9	−4 ± 7.8 (−20.2 to 5.9)	0.208
LV, mL	267.4 ± 254.9	15.1% ± 12.5%	1676.3 ± 265.2	1523.6 ± 268.3	35.5 ± 103.6 (−106 to 149)	0.313

Numbers are given as mean ± SD; numbers in parenthesis are ranges. *P* values represent differences between automated and manual measurements. Group A, PDFF < 10%; group B, PDFF ≥ 10%; group C, R2* ≥ 100 s^{-1}; group D, status post-hemihepatectomy.

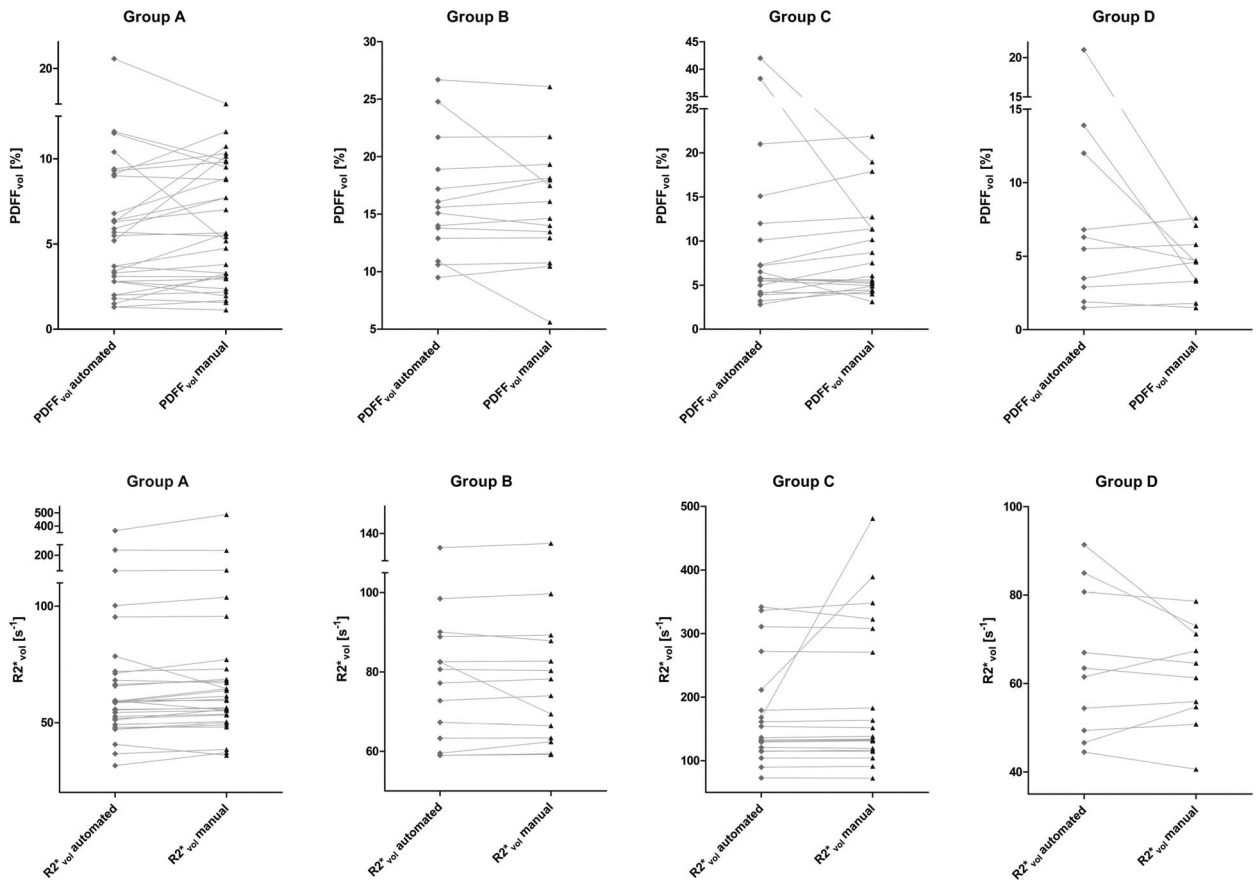

FIGURE 2. Case-by-case absolute differences in automated and manual $PDFF_{vol}$ (upper row) and $R2^*_{vol}$ measurements (lower row).

TABLE 4. Results of Linear Regression Analysis of Each Group

	$PDFF_{vol}$			$R2^*_{vol}$		
	Intercept	Slope	R^2	Intercept	Slope	R^2
All patients						
Group A (n = 30)	−0.261 (−1.825 to 1.302)	0.987 (0.770 to 1.205)	0.756	14.691 (9.239 to 20.144)	0.760 (0.713 to 0.806)	0.975
Group B (n = 14)	2.375 (−2.591 to 7.341)	0.889 (0.586 to 1.192)	0.773	2.952 (−6.452 to 12.356)	0.969 (0.854 to 1.084)	0.966
Group C (n = 20)	−2.123 (−9.231 to 4.986)	1.455 (0.759 to 2.152)	0.517	70.130 (16.311 to 123.948)	0.515 (0.276 to 0.754)	0.532
Group D (n = 10)	0.050 (−10.181 to 10.280)	1.682 (−0.429 to 3.792)	0.297	−14.607 (−48.577 to 19.363)	1.279 (0.737 to 1.820)	0.788
Patients with sampled volume partly outside the liver						
Group A (n = 11/30; 37%)	−0.168 (−4.179 to 3.843)	1.123 (0.656 to 1.590)	0.766	17.830 (9.235 to 26.425)	0.720 (0.667 to 0.773)	0.991
Group B (n = 5/14; 35%)	6.298 (−9.800 to 22.397)	0.744 (−0.246 to 1.734)	0.656	10.599 (−21.015 to 42.231)	0.905 (0.560 to 1.251)	0.959
Group C (n = 7/20; 35%)	2.253 (−24.477 to 28.984)	1.425 (−0.580 to 3.431)	0.400	128.843 (22.939 to 234.747)	0.183 (−0.198 to 0.565)	0.234
Group D (n = 8/10; 80%)	−3.224 (−18.269 to 11.820)	2.608 (−0.605 to 5.822)	0.397	−9.580 (−49.052 to 29.892)	1.216 (0.598 to 1.834)	0.794

Numbers in parenthesis are 95% confidence interval.
Group A, PDFF < 10%; group B, PDFF ≥ 10%; group C, R2* ≥ 100 s^{-1}; group D, status post-hemihepatectomy.

175.8 ± 212.5 mL (11.5% ± 12.0% of the whole inline-segmented volume [range, 0.6%–43.5%]).

The value of $PDFF_{vol}$ was similar between automated and manual measurements ($P = 0.154$), with a mean absolute difference of −0.3% ± 2.1% (range, −5.0% to 5.2%). The values of $R2^*_{vol}$ were significantly lower with automated measurements than with manual measurements ($P = 0.004$), with a mean absolute difference of −5.1 ± 22.7 s^{-1} (range, −123.4 to 13.8 s^{-1}; Fig. 5).

In patients with inline segmentation partly outside the liver, no significant difference between automated and manually measured $PDFF_{vol}$ and $R2^*_{vol}$ ($P = 0.413$ and $P = 0.175$, respectively) was seen. However, a tendency toward overestimation of $PDFF_{vol}$ with overestimation beyond the 95% limits of agreement in 2 of these 11 cases and toward underestimation of $R2^*_{vol}$ with underestimation beyond the 95% limits of agreement in 1 of these 11 cases was seen with automated measurements (Fig. 5).

The mean absolute difference between automated and manual measurements in patients with correct inline segmentation was small for $PDFF_{vol}$ with 1% ± 1.2% and for R2* with 1.4 ± 2.4 s^{-1}.

Group B (PDFF ≥ 10%)

With inline liver segmentation, in 5 (35%) of 14 cases, the sampled volume was located partly outside the liver on multiecho results. The mean volume located outside the liver was 232.9 ± 170.9 mL (11.9% ± 8.4% of the whole inline-segmented volume [range, 4.2%–25.8%]).

The values of $PDFF_{vol}$ and $R2^*_{vol}$ with automated and manual measurements were similar ($P = 0.843$ and $P = 0.358$, respectively). For $PDFF_{vol,}$ the mean absolute difference was 0.6% ± 2.5% (range, −1.9% to 7.3%). For $R2^*_{vol}$, the mean absolute difference was 0.5 ± 3.8 s^{-1} (range, −2.9 to 13.1 s^{-1}; Fig. 5).

In patients with inline segmentation partly outside the liver, no significant difference between automated and manually measured $PDFF_{vol}$ and $R2^*_{vol}$ was seen ($P = 0.438$ and $P = 1.000$, respectively). A tendency toward overestimation of $PDFF_{vol}$ and $R2^*_{vol}$ was seen with overestimation beyond the 95% limits of agreement in 1 of these 5 cases with automated measurements (Fig. 5).

The mean absolute difference between automated and manual measurements in patients with correct inline segmentation was small for $PDFF_{vol}$ with 0.3% ± 0.8% and for R2* with 0.5 ± 1.4 s^{-1}.

Group C ($R2^*_{vol}$ > 100 s^{-1})

With inline liver segmentation, in 7 (35%) of 20 cases, the sampled volume was located partly outside the liver on multiecho results. The mean volume located outside the liver was 436 ± 599.8 mL (27.7% ± 39.1% of the whole inline-segmented volume [range, 1.6%–100%]; Fig. 6).

The values of $PDFF_{vol}$ and $R2^*_{vol}$ with automated and manual measurements were similar ($P = 0.330$ and $P = 0.123$, respectively). For $PDFF_{vol}$, the mean absolute difference was 1.8% ± 8.1% (range, −2.9% to 27.0%). For $R2^*_{vol}$, the mean absolute difference was −24.6 ± 78.8 s^{-1} (range, −313.0 to 19.3 s^{-1}; Fig. 5).

In patients with inline segmentation partly outside the liver, no significant difference between automated and manual measured $PDFF_{vol}$ and $R2^*_{vol}$ was seen ($P = 0.398$ and $P = 0.091$, respectively). A tendency toward overestimation of $PDFF_{vol}$ was seen with overestimation beyond the 95% limits of agreement in 2 of these 7 cases and toward underestimation of $R2^*_{vol}$ with underestimation beyond the 95% limits of agreement in 1 of these 7 cases with automated measurements (Fig. 5).

The mean absolute difference between automated and manual measurements in patients with correct inline segmentation was small for $PDFF_{vol}$ with 1.1% ± 1.2% and for R2* with −0.2 ± 6.5 s^{-1}.

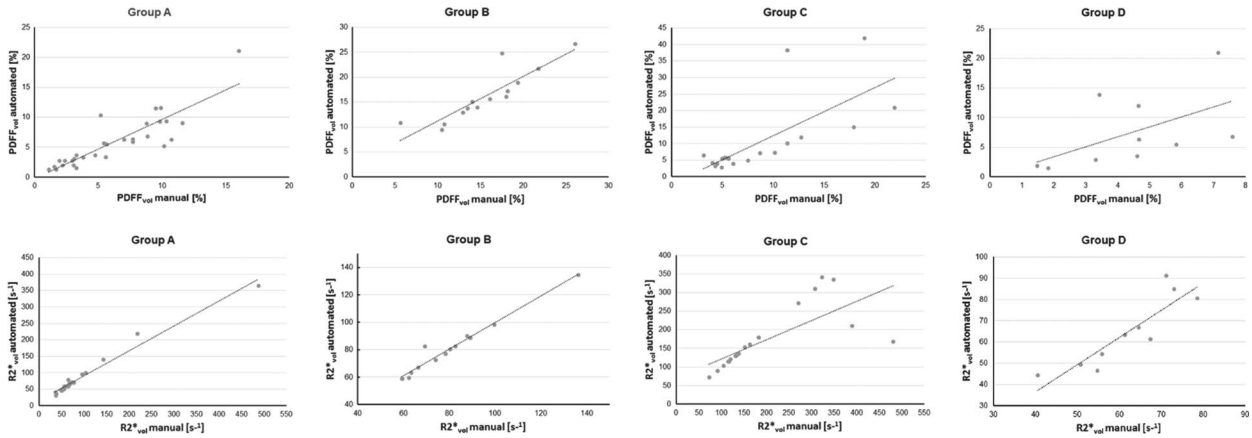

FIGURE 3. Scatter plots of PDFF$_{vol}$ and R2*$_{vol}$ showing a correlation between automated and manual measurements. The dashed lines represent the reference line.

Group D (Status Post-Hemihepatectomy)

With inline liver segmentation, in 8 (80%) of 10 cases, the sampled volume was located partly outside the liver on multiecho results. The mean volume located outside the liver was 267.4 ± 272.5 mL (15.4% ± 13.3% of the whole inline-segmented volume [range, 0.7%–42.1%]; Fig. 7).

The values of PDFF$_{vol}$ and R2*$_{vol}$ with automated and manual measurements were similar ($P = 0.322$ and $P = 0.375$, respectively). For PDFF$_{vol}$, the mean absolute difference was 3.1% ± 5.4% (range, −1.1% to 13.9%) with a mean relative difference of 65.3% ± 114.1% (range, −23.9% to 306.2%). For R2*$_{vol}$, the mean absolute difference was 2.6 ± 8.3 s^{-1} (range, −8.1 to 20.2 s^{-1}) with a mean relative difference of 3.6% ± 12.4% (range, −14.8% to 28.3%; Fig. 5).

In patients with inline segmentation partly outside the liver, no significant difference between automated and manually measured PDFF$_{vol}$ and R2*$_{vol}$ was seen ($P = 0.313$ and $P = 0.208$, respectively). A tendency toward overestimation of PDFF$_{vol}$ and R2*$_{vol}$ was seen with overestimation beyond the 95% limits of agreement in 1 of these 8 cases with automated measurements (Fig. 5).

The mean absolute difference between automated and manual measurements in patients with correct inline segmentation was small for PDFF$_{vol}$ with 0.2% ± 0.6% and for R2* with 2.9 ± 5.2 s^{-1}.

DISCUSSION

In our study, the automated inline LV segmentation propagated from dual-echo Dixon images to multiecho Dixon images performed well in patients without and with hepatic steatosis, where on average 11.5% to 11.9% of the segmented volume was located outside the liver in 35% and 36.7% of cases. This inline segmentation produced similar automated PDFF$_{vol}$ and R2*$_{vol}$ measurements in the liver compared with manual measurements. In cases with iron overload, extrahepatic areas were erroneously included to a greater extent (on average, 27% of the segmented volume located outside the liver), and in cases with post-hemihepatectomy, they were erroneously included more frequently (80% of the cases) by the inline segmentation algorithm. In the latter, a tendency toward overestimation of PDFF$_{vol}$ was seen with the automated measurements compared with manual measurements with the greatest individual differences in automated and manual PDFF$_{vol}$. However, these differences did not reach significance for any PDFF$_{vol}$ measurements or most R2*$_{vol}$ measurements and were relatively small in magnitude. Overall, the grading of fat deposition was concordant between automated and manual PDFF$_{vol}$ in 59 (80%) of 74 cases. In only 2 cases, an abnormal fat deposition of greater than 5% PDFF was missed with automated measurements.

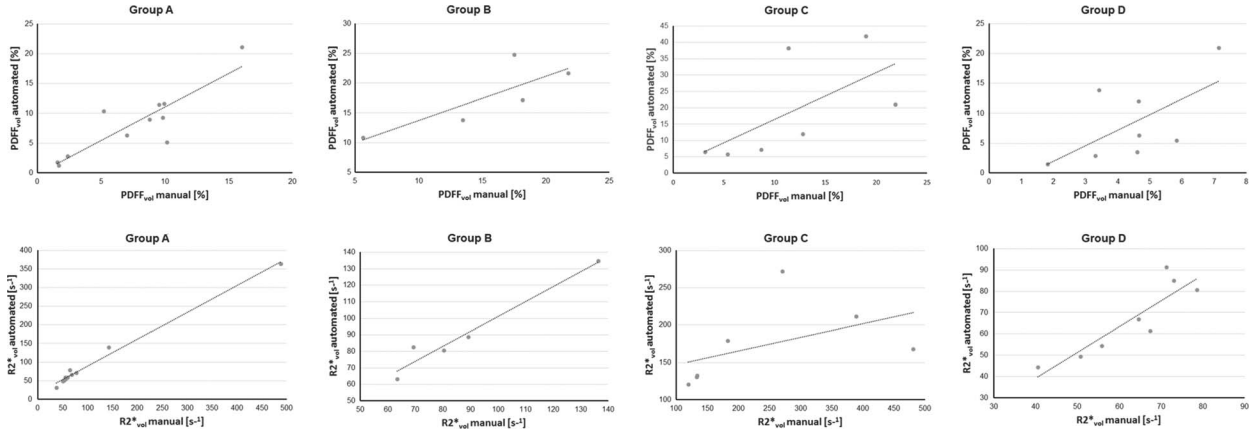

FIGURE 4. Scatter plots of PDFF$_{vol}$ and R2*$_{vol}$ showing a correlation between automated and manual measurements in patients where the sample volume was located partly outside the liver. The dashed lines represent the reference line.

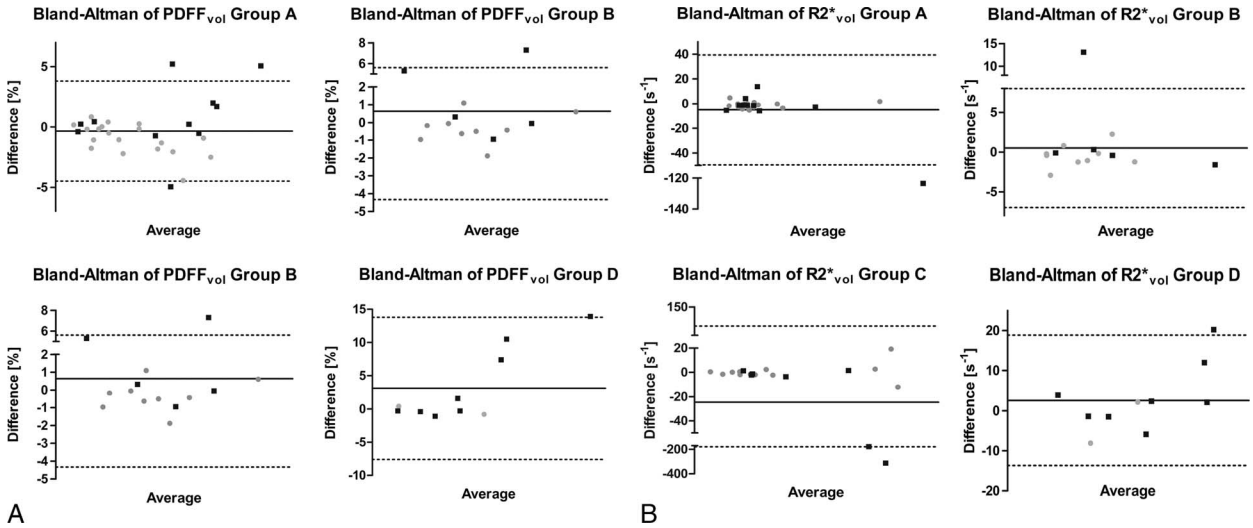

FIGURE 5. Bland-Altman plots of PDFF$_{vol}$ (A) and R2*$_{vol}$ (B) showing absolute differences between automated and manual measurements. Gray dots represent cases with adequate segmentation of the liver parenchyma. Black squares represent cases with segmentation partly outside the liver parenchyma. The dashed lines represent the 95% limits of agreement.

The automated measurements come with the advantage of user independence and short user interaction times. In particular, the quantification of liver deposition disease in the whole LV requires outlining of the whole liver contour on several slices, when performed manually. This manual segmentation is a time-consuming procedure compared with semiautomated or completely automated software as demonstrated in previous studies and in many cases must be done at a workstation separate from the one used for interpretation of the rest of the MRI examination.[18,19] An automated algorithm for quantification of fat or iron content in the whole liver on multiecho Dixon sequences would facilitate the implementation of these measures into clinical practice.

Another approach to shorten postprocessing time would be to use a single ROI measurement in the liver. However, similarly to liver biopsy, this single ROI may not be representative because liver deposition disease is known to show inhomogeneous distributions within the liver.[12,20] Furthermore, when measuring liver deposition disease over time, the placement of a single ROI in the same location as that on previous examinations would require meticulous matching with previous images, and duplicating

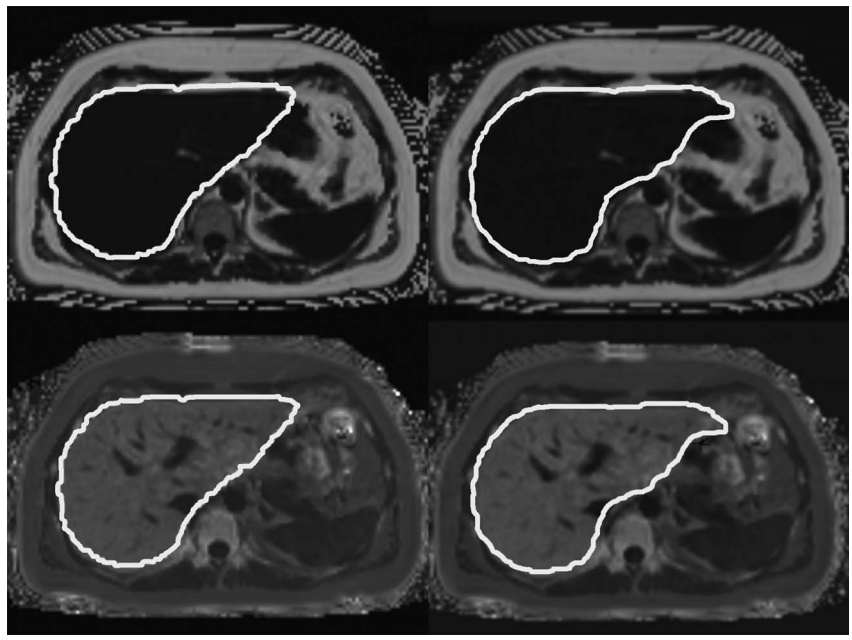

FIGURE 6. Example of a correctly segmented LV in a 42-year-old female patient in group C. Fat percentage maps (upper row) and R2* maps (lower row) of the multiecho Dixon sequence with automatic (left side) and manual (right side) LV segmentation. PDFF$_{vol}$ with automated measurement was 5.5%, and that with manual measurement was 5.6% (relative difference, 1.9%). R2*$_{vol}$ with automated measurement was 161.3 s^{-1}, and that with manual measurement was 162.5 s^{-1} (relative difference, 1.5%).

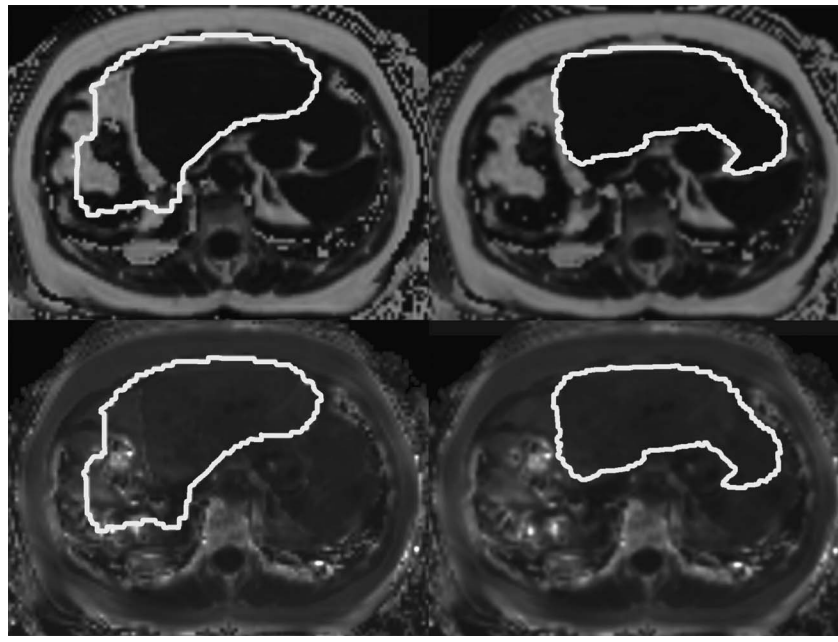

FIGURE 7. Example of mis-segmented LV in a 58-year-old female patient in group D. Fat percentage maps (upper row) and R2* maps (lower row) of the multiecho Dixon sequence with automatic (left side) and manual (right side) LV segmentation. $PDFF_{vol}$ with automated measurement was 21%, and that with manual measurement was 7.1% (relative difference, −66%). $R2^*_{vol}$ with automated measurement was 91.4 s^{-1}, and that with manual measurement was 71.2 s^{-1} (relative difference, −22.1%). The mis-segmentation on multiecho results was probably due to the abnormal liver shape after major liver resection.

multiple ROI placements would be even more time consuming.[21] Currently, it is recommended that multiple ROIs or large ROIs tailored to exclude the blood vessels be used for measurement.[22]

The automated $PDFF_{vol}$ and $R2^*_{vol}$ measurements could be part of a fast and standardized MRI protocol as proposed by Hetterich et al[23] for quantitative evaluation of liver fat and iron deposition in clinical or scientific examinations. Furthermore, the automated algorithm may allow for more repeatable and comparable measurements between examinations, which would be of advantage when hepatic steatosis and iron content were measured on serial follow-up examinations, although we did not test that hypothesis in the current work.

For example, in the setting of preoperative workup, a simple and noninvasive method for liver fat quantification in patients before surgery can provide important information for the surgeon and improve patient's outcome. d'Assignies et al[24] showed that liver steatosis is an independent risk factor for severe postoperative complications after major hepatic resection. An MRI-based liver fat quantification could, first, serve for detection of liver steatosis in this preoperative setting and, second, in case of a preoperative attempt to reduce the degree of liver steatosis and therefore the risk of peritoperative and postoperative complications (ie, short-term caloric restriction, omega-3 fatty acid supplement).[25,26]

Another clinical application might be the quantification and monitoring of fatty liver disease, such as nonalcoholic fatty liver diseases as the most prevalent metabolic liver disease worldwide with a prevalence of up to 45%.[27,28] In the light of recent drug developments for the treatment of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis, a method for straightforward and preferably automated liver fat quantification such as the proposed algorithm of automated inline liver segmentation is desirable. In a recent study on intraexamination and interexamination repeatability of PDFF measurements, a longitudinal change in PDFF greater than 1.8% was considered a real change.[29] Considering this as a reference, only the measurements with correct outline of the LV of our study are within this limit and should be considered. In cases with incorrect measurements outside the liver parenchyma, a refinement of the inline liver segmentation algorithm may be warranted to obtain automated results. Regarding the diagnostic accuracy of PDFF, previous studies have shown a high correlation of volumetric PDFF measurements with MR spectroscopy.[6]

The liver segmentation algorithm used in this study takes as input water-only image volumes from a routinely used T1-weighted dual-echo Dixon VIBE sequence. It was trained on a collection of patient and volunteer data from an unspecific cohort. Because of the inline implementation, without the possibility for manual correction, the segmentation result needs to be verified, and subsequent measurement results need to be disregarded in case of failure. Possible causes for incorrect segmentation results are deviations in image contrast, signal intensity, and patient pathology, relative to the original training data set. In particular, inline segmentation results in group C were very likely confounded by signal loss from iron overload.

Another reason for incorrect inline LV segmentation seems to be an abnormal liver shape after major liver resection. In our study, we saw a distinctly higher frequency of inline-segmented volume partly outside the liver in group D with status post-hemihepatectomy (80%) compared with groups A, B, and C (36.7%, 35%, and 35%). The explanation for this missegmentation in the hemihepatectomy group is that the inline liver segmentation algorithm has been trained on normally shaped livers. Furthermore, it is assumed that also any other cause of an abnormal liver shape, such as cirrhosis or wedge resection, which was present in 22% and 9% of our 23 cases with mis-segmentation, may lead to failure of the inline liver segmentation.

A side product of an automated volumetric analysis could be the LV itself. Liver deposition disease and LV are 2 relevant factors in planning of major liver resections because they both influence postoperative outcome and could be ideally assessed in one

MRI protocol.[30] The inline segmentation algorithm used in our study was not designed for quantitative volume measurements, and we found significantly smaller LVs derived from inline segmentation compared with manual segmentation.

The following study limitations must be taken into account. First, in this study population, we limited the number of patients with nonsteatotic and nonsiderotic livers, so diseased livers are overrepresented in the total study population. Therefore, the performance of the inline algorithm in this study may not be representative of the performance in a clinical population with a preponderance of normal livers. Second, we did not correlate our $PDFF_{vol}$ and $R2^*_{vol}$ measurements with histopathology or spectroscopy as the standard of reference, because the assessment of diagnostic accuracy of the sequence was not within the scope of our study and has been validated previously.[7,23,31] Third, we assigned the patients to the different study groups on the basis of single ROI measurements of PDFF and R2* from the clinical reports. As we analyzed PDFF and R2* values from whole liver segmentations in the study, which are potentially different from single ROI measurements, some $PDFF_{vol}$ measurements were greater than 10% (4/30 with automated measurements and 5/30 with manual measurements) in group A and greater than 10% (1/14 with automated and manual measurements) in group B, and $R2^*_{vol}$ measurements in group C were greater than 100 s^{-1} (2/20 with automated and manual measurements).

In conclusion, the inline LV segmentation performed well in patients without and with hepatic steatosis with similar automated $PDFF_{vol}$ and $R2^*_{vol}$ measurements in the liver compared with manual measurements. In cases with iron overload and post-hemihepatectomy, extrahepatic areas were erroneously included to a greater extent by the inline segmentation algorithm, with a tendency toward overestimation of $PDFF_{vol}$ compared with manual measurements. Especially for these cases with an abnormal liver contour and with severely abnormal liver signal intensity as in iron overload, a refinement of the inline liver segmentation algorithm may be warranted. However, an automated workflow to quantify fat and iron in the liver could be an important tool, easy to implement in daily workflows, to detect or monitor metabolic liver diseases.

REFERENCES

1. Alustiza JM, Artetxe J, Castiella A, et al. MR quantification of hepatic iron concentration. *Radiology*. 2004;230:479–484.
2. Gandon Y, Olivié D, Guyader D, et al. Non-invasive assessment of hepatic iron stores by MRI. *Lancet*. 2004;363:357–362.
3. Idilman IS, Aniktar H, Idilman R, et al. Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. *Radiology*. 2013;267:767–775.
4. Permutt Z, Le TA, Peterson MR, et al. Correlation between liver histology and novel magnetic resonance imaging in adult patients with non-alcoholic fatty liver disease—MRI accurately quantifies hepatic steatosis in NAFLD. *Aliment Pharmacol Ther*. 2012;36:22–29.
5. Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. *J Magn Reson Imaging*. 2012;36:1011–1014.
6. Henninger B, Zoller H, Kannengiesser S, et al. 3D multiecho Dixon for the evaluation of hepatic iron and fat in a clinical setting. *J Magn Reson Imaging*. 2017;46:793–800.
7. Zhong X, Nickel MD, Kannengiesser SA, et al. Liver fat quantification using a multi-step adaptive fitting approach with multi-echo GRE imaging. *Magn Reson Med*. 2014;72:1353–1365.
8. Bashir MR, Zhong X, Nickel MD, et al. Quantification of hepatic steatosis with a multistep adaptive fitting MRI approach: prospective validation against MR spectroscopy. *AJR Am J Roentgenol*. 2015;204:297–306.
9. McGill DB, Rakela J, Zinsmeister AR, et al. A 21-year experience with major hemorrhage after percutaneous liver biopsy. *Gastroenterology*. 1990;99:1396–1400.
10. Rockey DC, Caldwell SH, Goodman ZD, et al. Liver biopsy. *Hepatology*. 2009;49:1017–1044.
11. El-Badry AM, Breitenstein S, Jochum W, et al. Assessment of hepatic steatosis by expert pathologists: the end of a gold standard. *Ann Surg*. 2009;250:691–697.
12. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology*. 2005;128:1898–1906.
13. Bashir MR, Zhong X, Dale BM, et al. Automated patient-tailored screening of the liver for diffuse steatosis and iron overload using MRI. *AJR Am J Roentgenol*. 2013;201:583–588.
14. Bashir MR, Dale BM, Merkle EM, et al. Automated liver sampling using a gradient dual-echo Dixon-based technique. *Magn Reson Med*. 2012;67:1469–1477.
15. Kohlberger T, Sofka M, Zhang JD, et al. Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In: Fichtinger G, Martel A, Peters T, eds. *Medical Image Computing and Computer-Assisted Intervention*—MICCAI 2011. Berlin, Heidelberg: Springer; 2011;6893(pt III):338+.
16. Lay N, Birkbeck N, Zhang J, et al. Rapid multi-organ segmentation using context integration and discriminative models. *Inf Process Med Imaging*. 2013;23:450–462.
17. Ling HB, Zhou SK, Zheng YF, et al. *Hierarchical, Learning-Based Automatic Liver Segmentation*. 2008 I.E. Conference on Computer Vision and Pattern Recognition; Anchorage, AK; 2008;1–12:405+.
18. Nakayama Y, Li Q, Katsuragawa S, et al. Automated hepatic volumetry for living related liver transplantation at multisection CT. *Radiology*. 2006;240:743–748.
19. Suzuki K, Epstein ML, Kohlbrenner R, et al. Quantitative radiology: automated CT liver volumetry compared with interactive volumetry and manual volumetry. *AJR Am J Roentgenol*. 2011;197:W706–W712.
20. Bonekamp S, Tang A, Mashhood A, et al. Spatial distribution of MRI-determined hepatic proton density fat fraction in adults with nonalcoholic fatty liver disease. *J Magn Reson Imaging*. 2014;39:1525–1532.
21. Sofue K, Mileto A, Dale BM, et al. Interexamination repeatability and spatial heterogeneity of liver iron and fat quantification using MRI-based multistep adaptive fitting algorithm. *J Magn Reson Imaging*. 2015;42:1281–1290.
22. Vu KN, Gilbert G, Chalut M, et al. MRI-determined liver proton density fat fraction, with MRS validation: Comparison of regions of interest sampling methods in patients with type 2 diabetes. *J Magn Reson Imaging*. 2016;43:1090–1099.
23. Hetterich H, Bayerl C, Peters A, et al. Feasibility of a three-step magnetic resonance imaging approach for the assessment of hepatic steatosis in an asymptomatic study population. *Eur Radiol*. 2016;26:1895–1904.
24. d'Assignies G, Fayard C, Leitao H, et al. Liver steatosis assessed by preoperative MRI: an independent risk factor for severe complications after major hepatic resection. *Surgery*. 2016;159:1050–1057.
25. McCormack L, Petrowsky H, Jochum W, et al. Hepatic steatosis is a risk factor for postoperative complications after major hepatectomy: a matched case-control study. *Ann Surg*. 2007;245:923–930.
26. Reeves JG, Suriawinata AA, Ng DP, et al. Short-term preoperative diet modification reduces steatosis and blood loss in patients undergoing liver resection. *Surgery*. 2013;154:1031–1037.
27. Armstrong MJ, Adams LA, Canbay A, et al. Extrahepatic complications of nonalcoholic fatty liver disease. *Hepatology*. 2014;59:1174–1197.

28. Browning JD, Szczepaniak LS, Dobbins R, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology*. 2004;40:1387–1395.

29. Tyagi A, Yeganeh O, Levin Y, et al. Intra- and inter-examination repeatability of magnetic resonance spectroscopy, magnitude-based MRI, and complex-based MRI for estimation of hepatic proton density fat fraction in overweight and obese children and adults. *Abdom Imaging*. 2015;40:3070–3077.

30. d'Assignies G, Kauffmann C, Boulanger Y, et al. Simultaneous assessment of liver volume and whole liver fat content: a step towards one-stop shop preoperative MRI protocol. *Eur Radiol*. 2011;21: 301–309.

31. Tang A, Desai A, Hamilton G, et al. Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology*. 2015;274: 416–425.